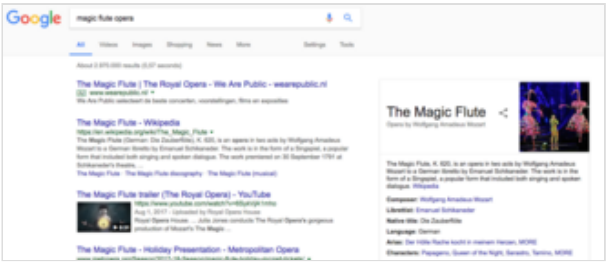
WIKIPEDIA 25
25 years of the free encyclopedia

Search engine

A **search engine** is a software system that provides hyperlinks to web pages, and other relevant information on the Web in response to a user's query. The user enters a query in a web browser or a mobile app, and the search results are typically presented as a list of hyperlinks accompanied by textual summaries and images. Users also have the option of limiting a search to specific types of results, such as images, videos, or news.



A Google search result for the phrase "magic flute opera"

For a search provider, its engine is part of a distributed computing system that can encompass many data centers throughout the world. The speed and accuracy of an engine's response to a query are based on a complex system of indexing that is continuously updated by automated web crawlers. This can include data mining the files and databases stored on web servers, although some content is not accessible to crawlers.

There have been many search engines since the dawn of the Web in the 1990s; however, Google Search became the dominant one in the 2000s and has remained so. As of May 2025, according to StatCounter, Google holds approximately 89–90 % of the worldwide search share, with competitors trailing far behind: Bing (~4 %), Yandex (~2.5 %), Yahoo! (~1.3 %), DuckDuckGo (~0.8 %), and Baidu (~0.7 %).[1] Notably, this marks the first time in over a decade that Google's share has fallen below the 90% threshold. The business of websites improving their visibility in search results, known as marketing and optimization, has thus largely focused on Google.

History

Pre-1990s

In 1945, Vannevar Bush described an information retrieval system that would allow a user to access a great expanse of information, all at a single desk, which he called a memex.[2] He described this system in an article titled "As We May Think" in *The Atlantic Monthly*.[3] The memex was intended to give a user the capability to overcome the ever-increasing difficulty of locating information in ever-growing centralized indices of scientific work. Vannevar Bush envisioned libraries of research with connected annotations, which are similar to modern hyperlinks.[4]

Link analysis eventually became a crucial component of search engines through algorithms such as Hyper Search and PageRank.[5][6]

1990s: Birth of search engines

The first internet search engines predate the debut of the Web in December 1990: WHOIS user search dates back to 1982,[7] and the Knowbot Information Service multi-network user search was first implemented in 1989.[8] The first well documented search engine that searched content files, namely FTP files, was Archie, which debuted on 10 September 1990.[9]

Prior to September 1993, the World Wide Web was entirely indexed by hand. There was a list of webservers edited by Tim Berners-Lee and hosted on the CERN webserver. One snapshot of the list in 1992 remains,[10] but as more and more web servers went online the central list could no longer keep up. On the NCSA site, new servers were announced under the title "What's New!".[11]

The first tool used for searching content (as opposed to users) on the Internet was Archie.[12] The name stands for "archive" without the "v".[13] It was created by Alan Emtage,[13][14][15][16] computer science student at McGill University in Montreal, Quebec, Canada. The program downloaded the directory listings of all the files located on public anonymous FTP (File Transfer Protocol) sites, creating a searchable database of file names; however, Archie Search Engine did not index the contents of these sites since the amount of data was so limited it could be readily searched manually.

The rise of Gopher (created in 1991 by Mark McCahill at the University of Minnesota) led to two new search programs, Veronica and Jughead. Like Archie, they searched the file names and titles stored in Gopher index systems. Veronica (Very Easy Rodent-Oriented Net-wide Index to Computerized Archives) provided a keyword search of most Gopher menu titles in the entire Gopher listings. Jughead (Jonzy's Universal Gopher Hierarchy Excavation And Display) was a tool for obtaining menu information from specific Gopher servers. While the name of the search engine "Archie Search Engine" was not a reference

Timeline (full list)

Year	Engine	Current status
1993	W3Catalog	Inactive
	ALIWEB	Inactive
	JumpStation	Inactive
	WWW Worm	Inactive
1994	WebCrawler	Active
	Go.com	Inactive, redirects to Disney
	Lycos	Active
	Infoseek	Inactive, redirects to Disney
1995	Yahoo! Search	Active, initially a search function for Yahoo! Directory
	Daum	Active
	Search.ch	Active
	Magellan	Inactive
	Excite	Active
	MetaCrawler	Active
	AltaVista	Inactive, acquired by Yahoo! in 2003, since 2013 redirects to Yahoo!
	SAPO	Active
1996	RankDex	Inactive, incorporated into Baidu in 2000
	Dogpile	Active
	HotBot	Inactive (used Inktomi search technology)
	Ask Jeeves	Active (rebranded ask.com)
1997	AOL NetFind	Active (rebranded AOL Search since 1999)
	goo.ne.jp	Active
	Northern Light	Inactive
	Yandex	Active
1998	Google	Active
	Ixquick	Active as Startpage.com
	MSN Search	Active as Bing
	empas	Inactive (merged with NATE)
1999	AlltheWeb	Inactive (URL redirected to Yahoo!)

to the Archie comic book series, "Veronica" and "Jughead" are characters in the series, thus referencing their predecessor.

In the summer of 1993, no search engine existed for the web, though numerous specialized catalogs were maintained by hand. Oscar Nierstrasz at the University of Geneva wrote a series of Perl scripts that periodically mirrored these pages and rewrote them into a standard format. This formed the basis for W3Catalog, the web's first primitive search engine, released on September 2, 1993.[17]

In June 1993, Matthew Gray, then at MIT, produced what was probably the first web robot, the Perl-based World Wide Web Wanderer, and used it to generate an index called "Wandex". The purpose of the Wanderer was to measure the size of the World Wide Web, which it did until late 1995. The web's second search engine Aliweb appeared in November 1993. Aliweb did not use a web robot, but instead depended on being notified by website administrators of the existence at each site of an index file in a particular format.

JumpStation (created in December 1993[18] by Jonathon Fletcher) used a web robot to find web pages and to build its index, and used a web form as the interface to its query program. It was thus the first WWW resource-discovery tool to combine the three essential features of a web search engine (crawling, indexing, and searching) as described below. Because of the limited resources available on the platform it ran on, its indexing and hence searching were limited to the titles and headings found in the web pages the crawler encountered.

One of the first "all text" crawler-based search engines was WebCrawler, which came out in 1994. Unlike its predecessors, it allowed users to search for any word in any web page, which has become the standard for all major search engines since. It was also the search engine that was widely known by the public. Also, in 1994, Lycos (which started at Carnegie Mellon University) was launched and became a major commercial endeavor.

	GenieKnows	Inactive, rebranded Yellowee (was redirecting to justlocalbusiness.com)
	Naver	Active
	Teoma	Inactive (redirect to Ask.com)
2000	Baidu	Active
	Exalead	Inactive
	Gigablast	Inactive
2001	Kartoo	Inactive
2003	Info.com	Active
2004	A9.com	Inactive
	Clusty	Active, Yippy, previously Clusty, now owns Togoda.com
	Mojeek	Active
	Sogou	Active
2005	SearchMe	Inactive
	KidzSearch	Active, Google Search
2006	Soso	Inactive, merged with Sogou
	Quaero	Inactive
	Search.com	Active
	ChaCha	Inactive
	Ask.com	Active
	Live Search	Active as Bing, rebranded MSN Search
2007	wikiseek	Inactive
	Sproose	Inactive
	Wikia Search	Inactive
	Blackle.com	Active, Google Search
2008	Powerset	Inactive (redirects to Bing)
	Picollator	Inactive
	Viewzi	Inactive
	Boogami	Inactive
	LeapFish	Inactive
	Forestle	Inactive (redirects to Ecosia)
	DuckDuckGo	Active
	TinEye	Active

The first popular search engine on the Web was Yahoo! Search.[19] The first product from Yahoo!, founded by Jerry Yang and David Filo in January 1994, was a Web directory called Yahoo! Directory. In 1995, a search function was added, allowing users to search Yahoo! Directory.[20][21] It became one of the most popular ways for people to find web pages of interest, but its search function operated on its web directory, rather than its full-text copies of web pages.

Soon after, a number of search engines appeared and vied for popularity. These included Magellan, Excite, Infoseek, Inktomi, Northern Light, and AltaVista. Information seekers could also browse the directory instead of doing a keyword-based search.

In 1996, Robin Li developed the RankDex site-scoring algorithm for search engines results page ranking[22][23][24] and received a US patent for the technology.[25] It was the first search engine that used hyperlinks to measure the quality of websites it was indexing,[26] predating the very similar algorithm patent filed by Google two years later in 1998.[27] Larry Page referenced Li's work in some of his U.S. patents for PageRank.[28] Li later used his RankDex technology for the Baidu search engine, which was founded by him in China and launched in 2000.

In 1996, Netscape was looking to give a single search engine an exclusive deal as the featured search engine on Netscape's web browser. There was so much interest that instead, Netscape struck deals with five of the major search engines: for $5 million a year, each search engine would be in rotation on the Netscape search engine page. The five engines were Yahoo!, Magellan, Lycos, Infoseek, and Excite.[29][30]

2009	Bing	Active, rebranded Live Search
	Yebol	Inactive
	Scout (Goby)	Active
	NATE	Active
	Ecosia	Active
	Startpage.com	Active, sister engine of Ixquick
2010	Blekko	Inactive, sold to IBM
	Cuil	Inactive
	Yandex (English)	Active
	Parsijoo	Active
2011	YaCy	Active, P2P
2012	Volunia	Inactive
2013	Qwant	Active
2014	Egerin	Active, Kurdish / Sorani
	Swisscows	Active
	Searx	Active
2015	Yooz	Inactive
	Cliqz	Inactive
2016	Kiddle	Active, Google Search
2017	Presearch	Active
2018	Kagi	Active
2020	Petal	Active
2021	Brave Search	Active
	You.com	Active
2022	Perplexity	Active

Google adopted the idea of selling search terms in 1998 from a small search engine company named goto.com. This move had a significant effect on the search engine business, which went from struggling to one of the most profitable businesses in the Internet.[31][32]

Search engines were also known as some of the brightest stars in the Internet investing frenzy that occurred in the late 1990s.[33] Several companies entered the market spectacularly, receiving record gains during their initial public offerings. Some have taken down their public search engine and are marketing enterprise-only editions, such as Northern Light. Many search engine companies were caught up in the dot-com bubble, a speculation-driven market boom that peaked in March 2000.

2000s–present: Post dot-com bubble

Around 2000, Google's search engine rose to prominence.[34] The company achieved better results for many searches with an algorithm called PageRank, as was explained in the paper *Anatomy of a Search Engine* written by Sergey Brin and Larry Page, the later founders of Google.[6] This iterative algorithm ranks web pages based on the number and PageRank of other web sites and pages that link there, on the premise that good or desirable pages are linked to more than others. Larry Page's patent for PageRank cites Robin Li's earlier RankDex patent as an influence.[28][24] Google also maintained a minimalist interface to its search engine. In contrast, many of its competitors embedded a search engine in a web portal. In fact, the Google search engine became so popular that spoof engines emerged such as Mystery Seeker.

By 2000, Yahoo! was providing search services based on Inktomi's search engine. Yahoo! acquired Inktomi in 2002, and Overture (which owned AlltheWeb and AltaVista) in 2003. Yahoo! switched to Google's search engine until 2004, when it launched its own search engine based on the combined technologies of its acquisitions.

Microsoft first launched MSN Search in the fall of 1998 using search results from Inktomi. In early 1999, the site began to display listings from Looksmart, blended with results from Inktomi. For a short time in 1999, MSN Search used results from AltaVista instead. In 2004, Microsoft began a transition to its own search technology, powered by its own web crawler (called msnbot).

Microsoft's rebranded search engine, Bing, was launched on June 1, 2009. On July 29, 2009, Yahoo! and Microsoft finalized a deal in which Yahoo! Search would be powered by Microsoft Bing technology.

As of 2019, active search engine crawlers include those of Baidu, Bing, Brave,[35] Google, DuckDuckGo, Gigablast, Mojeek, Sogou and Yandex.

Approach

A search engine maintains the following processes in near real time:[36]

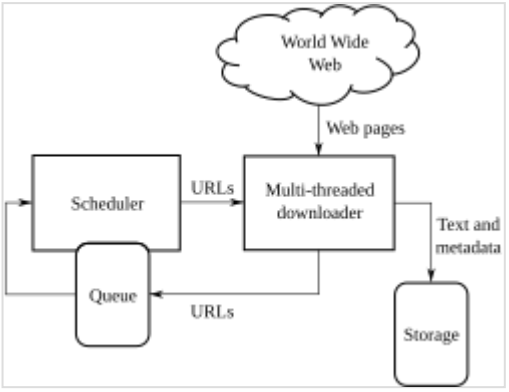
1. Web crawling
2. Indexing
3. Searching[37]

Web search engines get their information by web crawling from site to site. The "spider" checks for the standard filename *robots.txt*, addressed to it. The robots.txt file contains directives for search spiders, telling it which pages to crawl and which pages not to crawl. After checking for robots.txt and either finding it or not, the spider sends certain information back to be indexed depending on many factors, such as the titles, page content, JavaScript, Cascading Style Sheets (CSS), headings, or its metadata in HTML meta tags. After a certain number of pages crawled, amount of data indexed, or time spent on the website, the spider stops crawling and moves on. "[N]o web crawler may actually crawl the entire reachable web. Due to infinite websites, spider traps, spam, and other exigencies of the real web, crawlers instead apply a crawl policy to determine when the crawling of a site should be deemed sufficient. Some websites are crawled exhaustively, while others are crawled only partially".[38]

Indexing means associating words and other definable tokens found on web pages to their domain names and HTML-based fields. The associations are stored in a public database and accessible through web search queries. A query from a user can be a single word, multiple words or a sentence. The index helps find information relating to the query as quickly as possible.[37] Some of the techniques for indexing, and caching are trade secrets, whereas web crawling is a straightforward process of visiting all sites on a systematic basis.

Between visits by the *spider*, the cached version of the page (some or all the content needed to render it) stored in the search engine working memory is quickly sent to an inquirer. If a visit is overdue, the search engine can just act as a web proxy instead. In this case, the page may differ from the search terms indexed.[37] The cached page holds the appearance of the version whose words were previously indexed, so a cached version of a page can be useful to the website when the actual page has been lost, but this problem is also considered a mild form of linkrot.

Typically, when a user enters a query into a search engine it is a few keywords.[39] The index already has the names of the sites containing the keywords, and these are instantly obtained from the index. The real processing load is in generating the web pages that are the search results list: Every page in the entire list must be weighted according to information in the indexes.[37] Then the top search result item requires the lookup, reconstruction, and markup of the *snippets* showing the context of the keywords matched. These are only part of the processing each search results web page requires, and further pages (next to the top) require more of this post-processing.



High-level architecture of a standard Web crawler

Beyond simple keyword lookups, search engines offer their own GUI- or command-driven operators and search parameters to refine the search results. These provide the necessary controls for the user engaged in the feedback loop users create by *filtering* and *weighting* while refining the search results, given the initial pages of the first search results. For example, from 2007 the Google.com search engine has allowed one to *filter* by date by clicking "Show search tools" in the leftmost column of the initial search results page, and then selecting the desired date range.[40] It is also possible to *weight* by date because each page has a modification time. Most search engines support the use of the Boolean operators AND, OR and NOT to help end users refine the search query. Boolean operators are for literal searches that allow the user to refine and extend the terms of the search. The engine looks for the words or phrases exactly as entered. Some search engines provide an advanced feature called proximity search, which allows users to define the distance between keywords.[37] There is also concept-based searching where the research involves using statistical analysis on pages containing the words or phrases the user searches for.

The usefulness of a search engine depends on the relevance of the *result set* it gives back. While there may be millions of web pages that include a particular word or phrase, some pages may be more relevant, popular, or authoritative than others. Most search engines employ methods to rank the results to provide the "best" results first. How a search engine decides which pages are the best matches, and what order the results should be shown in, varies widely from one engine to another.[37] The methods also change over time as Internet usage changes and new techniques evolve. There are two main types of search engine

that have evolved: one is a system of predefined and hierarchically ordered keywords that humans have programmed extensively. The other is a system that generates an "inverted index" by analyzing texts it locates. This first form relies much more heavily on the computer itself to do the bulk of the work.

Most Web search engines are commercial ventures supported by advertising revenue and thus some of them allow advertisers to have their listings ranked higher in search results for a fee. Search engines that do not accept money for their search results make money by running search related ads alongside the regular search engine results. The search engines make money every time someone clicks on one of these ads.[41]

Local search

Local search is the process that optimizes the efforts of local businesses. They focus on ensuring consistent search results. It is important because many people determine where they plan to go and what to buy based on their searches.[42]

Market share

As of January 2022, Google is by far the world's most used search engine, with a market share of 90%, and the world's other most used search engines were Bing at 4%, Yandex at 2%, Yahoo! at 1%. Other search engines not listed have less than a 3% market share.[43] In 2024, Google's dominance was ruled an illegal monopoly in a case brought by the US Department of Justice.[44]

Russia and East Asia

As of late 2023 and early 2024, search engine market shares in Russia and East Asia have remained relatively stable but with some notable shifts due to geopolitical and technological developments.

In Russia, Yandex continues to dominate the search engine market with a share of approximately 70.7%, while Google holds around 23.3%.[45] Yandex also remains a key player in localized services including navigation, ride-hailing, and e-commerce, strengthening its ecosystem.

In China, Baidu remains the leading search engine with a market share of about 59.3% as of early 2024. Other domestic engines such as Sogou and 360 Search hold smaller shares. Google remains inaccessible in mainland China due to long-standing censorship issues, having exited the Chinese market in 2010 following disputes over censorship and cybersecurity.[46][47]

Bing, Microsoft's search engine, has maintained a niche presence in China with a market share around 13.6%, making it one of the few foreign search engines operating under local regulatory constraints.[48]

In South Korea, Naver continues to lead the domestic market, with a market share of 59.8%, followed by Google at 35.4% as of Q4 2023.[49] Naver’s strength lies in its localized services and integration with Korean content platforms.

In Japan, Google Japan currently holds the largest market share (around 76.2%), while Yahoo! Japan, operated by Z Holdings (a SoftBank and Naver joint venture), retains about 15.8% market share.[50]

In Taiwan, Google is the predominant search engine, commanding over 93% of the market, with Yahoo! Taiwan and Bing trailing far behind.[51]

Search engine bias

Although search engines are programmed to rank websites based on some combination of their popularity and relevancy, empirical studies indicate various political, economic, and social biases in the information they provide[52][53] and the underlying assumptions about the technology.[54] These biases can be a direct result of economic and commercial processes (e.g., companies that advertise with a search engine can become also more popular in its organic search results), and political processes (e.g., the removal of search results to comply with local laws).[55] For example, Google will not surface certain neo-Nazi websites in France and Germany, where Holocaust denial is illegal.

Biases can also be a result of social processes, as search engine algorithms are frequently designed to exclude non-normative viewpoints in favor of more "popular" results.[56] Indexing algorithms of major search engines skew towards coverage of U.S.-based sites, rather than websites from non-U.S. countries.[53]

Google Bombing is one example of an attempt to manipulate search results for political, social or commercial reasons.

Several scholars have studied the cultural changes triggered by search engines,[57] and the representation of certain controversial topics in their results, such as terrorism in Ireland,[58] climate change denial,[59] and conspiracy theories.[60]

Customized results and filter bubbles

There has been concern raised that search engines such as Google and Bing provide customized results based on the user's activity history, leading to what has been termed echo chambers or filter bubbles by Eli Pariser in 2011.[61] The argument is that search engines and social media platforms use algorithms to selectively guess what information a user would like to see, based on information about the user (such as location, past click behavior and search history). As a result, websites tend to show only information that agrees with the user's past viewpoint. According to Eli Pariser users get less exposure to conflicting viewpoints and are isolated intellectually in their own informational bubble. Since this problem has been identified, competing search engines have emerged that seek to avoid this problem by not tracking or "bubbling" users, such as DuckDuckGo. However many scholars have questioned Pariser's view, finding that there is little evidence for the filter bubble.[62][63][64] On the contrary, a number of studies trying to verify the existence of filter bubbles have found only minor levels of personalization in search,[64] that most people encounter a range of views when browsing online, and that Google news tends to promote mainstream established news outlets.[65][63]

Religious search engines

The global growth of the Internet and electronic media in the Arab and Muslim world during the last decade has encouraged Islamic adherents in the Middle East and Asian sub-continent, to attempt their own search engines, their own filtered search portals that would enable users to perform safe searches. More than usual *safe search* filters, these Islamic web portals categorizing websites into being either "halal" or "haram", based on interpretation of Sharia law. ImHalal came online in September 2011. Halalgoogling came online in July 2013. These use haram filters on the collections from Google and Bing (and others).[66]

While lack of investment and slow pace in technologies in the Muslim world has hindered progress and thwarted success of an Islamic search engine, targeting as the main consumers Islamic adherents, projects like Muxlim (a Muslim lifestyle site) received millions of dollars from investors like Rite Internet Ventures, and it also faltered. Other religion-oriented search engines are Jewogle, the Jewish version of Google,[67] and Christian search engine SeekFind.org. SeekFind filters sites that attack or degrade their faith.[68]

Search engine submission

Web search engine submission is a process in which a webmaster submits a website directly to a search engine. While search engine submission is sometimes presented as a way to promote a website, it generally is not necessary because the major search engines use web crawlers that will eventually find most web sites on the Internet without assistance. They can either submit one web page at a time, or they can submit the entire site using a sitemap, but it is normally only necessary to submit the home page of a web site as search engines are able to crawl a well designed website. There are two remaining reasons to submit a web site or web page to a search engine: to add an entirely new web site without waiting for a search engine to discover it, and to have a web site's record updated after a substantial redesign.

Some search engine submission software not only submits websites to multiple search engines, but also adds links to websites from their own pages. This could appear helpful in increasing a website's ranking, because external links are one of the most important factors determining a website's ranking. However, John Mueller of Google has stated that this "can lead to a tremendous number of unnatural links for your site" with a negative impact on site ranking.[69]

Comparison to social bookmarking

In comparison to search engines, a social bookmarking system has several advantages over traditional automated resource location and classification software, such as search engine spiders. All tag-based classification of Internet resources (such as web sites) is done by human beings, who understand the content of the resource, as opposed to software, which algorithmically attempts to determine the meaning and quality of a resource. Also, people can find and bookmark web pages that have not yet been noticed or indexed by web spiders.[70] Additionally, a social bookmarking system can rank a resource based on how many times it has been bookmarked by users, which may be a more useful metric for end-users than

systems that rank resources based on the number of external links pointing to it. However, both types of ranking are vulnerable to fraud, (see Gaming the system), and both need technical countermeasures to try to deal with this.

Technology

Archie

The first web search engine was Archie, created in 1990[71] by Alan Emtage, a student at McGill University in Montreal.

The primary method of storing and retrieving files was via the File Transfer Protocol (FTP). This was (and still is) a system that specified a common way for computers to exchange files over the Internet. It works like this: An administrator decides that they want to make files available from their computer. They set up a program on their computer, called an FTP server. When someone on the Internet wants to retrieve a file from this computer, they connect to it via another program called an FTP client. Any FTP client program can connect with any FTP server program as long as the client and server programs both fully follow the specifications set forth in the FTP protocol.

Initially, anyone who wanted to share a file had to set up an FTP server in order to make the file available to others. Later, "anonymous" FTP sites became repositories for files, allowing all users to post and retrieve them.

Even with archive sites, many important files were still scattered on small FTP servers. These files could be located only by the Internet equivalent of word of mouth: Somebody would post an e-mail to a message list or a discussion forum announcing the availability of a file.

Archie changed all that. It combined a script-based data gatherer, which fetched site listings of anonymous FTP files, with a regular expression matcher for retrieving file names matching a user query. (4) In other words, Archie's gatherer scoured FTP sites across the Internet and indexed all of the files it found. Its regular expression matcher provided users with access to its database.[72]

Veronica

In 1993, the University of Nevada System Computing Services group developed Veronica.[71] It was created as a type of searching device similar to Archie but for Gopher files. Another Gopher search service, called Jughead, appeared a little later, probably for the sole purpose of rounding out the comic-strip triumvirate. Jughead is an acronym for Jonzy's Universal Gopher Hierarchy Excavation and Display, although, like Veronica, it is probably safe to assume that the creator backed into the acronym. Jughead's functionality was pretty much identical to Veronica's, although it appears to be a little rougher around the edges.[72]

The Lone Wanderer

The World Wide Web Wanderer, developed by Matthew Gray in 1993[73] was the first robot on the Web and was designed to track the Web's growth. Initially, the Wanderer counted only Web servers, but shortly after its introduction, it started to capture URLs as it went along. The database of captured URLs became the Wandex, the first web database.

Matthew Gray's Wanderer created quite a controversy at the time, partially because early versions of the software ran rampant through the Net and caused a noticeable netwide performance degradation. This degradation occurred because the Wanderer would access the same page hundreds of times a day. The Wanderer soon amended its ways, but the controversy over whether robots were good or bad for the Internet remained.

In response to the Wanderer, Martijn Koster created Archie-Like Indexing of the Web, or ALIWEB, in October 1993. As the name implies, ALIWEB was the HTTP equivalent of Archie, and because of this, it is still unique in many ways.

ALIWEB does not have a web-searching robot. Instead, webmasters of participating sites post their own index information for each page they want listed. The advantage to this method is that users get to describe their own site, and a robot does not run about eating up Net bandwidth. The disadvantages of ALIWEB are more of a problem today. The primary disadvantage is that a special indexing file must be submitted. Most users do not understand how to create such a file, and therefore they do not submit their pages. This leads to a relatively small database, which meant that users are less likely to search ALIWEB than one of the large bot-based sites. This Catch-22 has been somewhat offset by incorporating other databases into the ALIWEB search, but it still does not have the mass appeal of search engines such as Yahoo! or Lycos.[72]

Excite

Excite, initially called Architext, was started by six Stanford undergraduates in February 1993. Their idea was to use statistical analysis of word relationships in order to provide more efficient searches through the large amount of information on the Internet. Their project was fully funded by mid-1993. Once funding was secured. they released a version of their search software for webmasters to use on their own web sites. At the time, the software was called Architext, but it now goes by the name of Excite for Web Servers.[72]

Excite was the first serious commercial search engine which launched in 1995.[74] It was developed in Stanford and was purchased for $6.5 billion by @Home. In 2001 Excite and @Home went bankrupt and InfoSpace bought Excite for $10 million.

Some of the first analysis of web searching was conducted on search logs from Excite[75][39]

Yahoo!

In April 1994, two Stanford University Ph.D. candidates, David Filo and Jerry Yang, created some pages that became rather popular. They called the collection of pages Yahoo! Their official explanation for the name choice was that they considered themselves to be a pair of yahoos.

As the number of links grew and their pages began to receive thousands of hits a day, the team created ways to better organize the data. In order to aid in data retrieval, Yahoo! (www.yahoo.com) became a searchable directory. The search feature was a simple database search engine. Because Yahoo! entries were entered and categorized manually, Yahoo! was not really classified as a search engine. Instead, it was generally considered to be a searchable directory. Yahoo! has since automated some aspects of the gathering and classification process, blurring the distinction between engine and directory.

The Wanderer captured only URLs, which made it difficult to find things that were not explicitly described by their URL. Because URLs are rather cryptic to begin with, this did not help the average user. Searching Yahoo! or the Galaxy was much more effective because they contained additional descriptive information about the indexed sites.

Lycos

At Carnegie Mellon University during July 1994, Michael Mauldin, on leave from CMU, developed the Lycos search engine.

Types of web search engines

Search engines on the web are sites enriched with facility to search the content stored on other sites. There is difference in the way various search engines work, but they all perform three basic tasks.[76]

1. Finding and selecting full or partial content based on the keywords provided.
2. Maintaining index of the content and referencing to the location they find
3. Allowing users to look for words or combinations of words found in that index.

The process begins when a user enters a query statement into the system through the interface provided.

Type	Example	Description
Conventional	librarycatalog	Search by keyword, title, author, etc.
Text-based	Google, Bing, Yahoo!	Search by keywords. Limited search using queries in natural language.
Voice-based	Google, Bing, Yahoo!	Search by keywords. Limited search using queries in natural language.
Multimedia search	QBIC, WebSeek, SaFe	Search by visual appearance (shapes, colors,..)
Q/A	Stack Exchange, NSIR	Search in (restricted) natural language
Clustering Systems	Vivisimo, Clusty, Togoda	
Research Systems	Lemur, Nutch	

There are basically three types of search engines: Those that are powered by robots (called crawlers; ants or spiders), those that are powered by human submissions, and those that are a hybrid of the two.

Crawler-based search engines are those that use automated software agents (called crawlers) that visit a Web site, read the information on the actual site, read the site's meta tags and also follow the links that the site connects to performing indexing on all linked Web sites as well. The crawler returns all that

information back to a central depository, where the data is indexed. The crawler will periodically return to the sites to check for any information that has changed. The frequency with which this happens is determined by the administrators of the search engine.

Human-powered search engines rely on humans to submit information that is subsequently indexed and catalogued. Only information that is submitted is put into the index.

In both cases, when a user queries a search engine to locate information, they're actually searching through the index that the search engine has created —they are not actually searching the Web. These indices are giant databases of information that is collected and stored and subsequently searched. This explains why sometimes a search on a commercial search engine, such as Yahoo! or Google, will return results that are, in fact, dead links. Since the search results are based on the index, if the index has not been updated since a Web page became invalid the search engine treats the page as still an active link even though it no longer is. It will remain that way until the index is updated.

So why will the same search on different search engines produce different results? Part of the answer to that question is because not all indices are going to be exactly the same. It depends on what the spiders find or what the humans submitted. But more important, not every search engine uses the same algorithm to search through the indices. The algorithm is what the search engines use to determine the relevance of the information in the index to what the user is searching for.

One of the elements that a search engine algorithm scans for is the frequency and location of keywords on a Web page. Those with higher frequency are typically considered more relevant. But search engine technology is becoming sophisticated in its attempt to discourage what is known as keyword stuffing, or spamdexing.

Another common element that algorithms analyze is the way that pages link to other pages in the Web. By analyzing how pages link to each other, an engine can both determine what a page is about (if the keywords of the linked pages are similar to the keywords on the original page) and whether that page is considered "important" and deserving of a boost in ranking. Just as the technology is becoming increasingly sophisticated to ignore keyword stuffing, it is also becoming more savvy to Web masters who build artificial links into their sites in order to build an artificial ranking.

Modern web search engines are highly intricate software systems that employ technology that has evolved over the years. There are a number of sub-categories of search engine software that are separately applicable to specific 'browsing' needs. These include web search engines (e.g. Google), database or structured data search engines (e.g. Dieselpoint), and mixed search engines or enterprise search. The more prevalent search engines, such as Google and Yahoo!, utilize hundreds of thousands computers to process trillions of web pages in order to return fairly well-aimed results. Due to this high volume of queries and text processing, the software is required to run in a highly dispersed environment with a high degree of superfluity.

Another category of search engines is scientific search engines. These are search engines which search scientific literature. The most prominent example is Google Scholar. Researchers are working on improving search engine technology by making them understand the content element of the articles, such as extracting theoretical constructs or key research findings.[77]

See also

- Comparison of web search engines
- Content moderation
- Filter bubble
- Google effect
- Information retrieval
- Use of web search engines in libraries
- Itpints
- List of search engines
- List of academic databases and search engines
- Question answering
- Search engine manipulation effect
- Search engine privacy
- Semantic Web
- Spell checker
- Timeline of web search engines
- Web development tools
- Web query
- Wikipedia:Search engine test, for a tutorial on using search engines for researching Wikipedia articles

References

1. StatCounter Global Stats – Search Engine Market Share (https://gs.statcounter.com/search-engine-market-share) (May 2025)
2. Bush, Vannevar (1 July 1945). "As We May Think" (https://web.archive.org/web/20120822132632/http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/4/). *The Atlantic*. Archived from the original (https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/) on 22 August 2012. Retrieved 22 February 2024.
3. "Search Engine History.com" (http://www.searchenginehistory.com/). *www.searchenginehistory.com*. Retrieved 2 July 2020.
4. "Penn State WebAccess Secure Login" (https://web.archive.org/web/20220122194212/https://webaccess.psu.edu/?cosign-scripts.libraries.psu.edu&https%3A%2F%2Fscripts.libraries.psu.edu%2Fscripts%2Fezproxyauth.php%3Furl=ezp.2aHR0cHM6Ly9pZWVleHBsb3JlLmllZWUub3JnL3N0YW1wL3N0YW1wLmpzcD90cD0mYXJudW1iZXI9ODUwOTU1MA--). *webaccess.psu.edu*. Archived from the original (https://webaccess.psu.edu/?cosign-scripts.libraries.psu.edu&https%3A%2F%2Fscripts.libraries.psu.edu%2Fscripts%2Fezproxyauth.php%3Furl=ezp.2aHR0cHM6Ly9pZWVleHBsb3JlLmllZWUub3JnL3N0YW1wL3N0YW1wLmpzcD90cD0mYXJudW1iZXI9ODUwOTU1MA--) on 22 January 2022. Retrieved 2 July 2020.
5. Marchiori, Massimo (1997). "The Quest for Correct Information on the Web: Hyper Search Engines" (http://www.w3.org/People/Massimo/papers/WWW6/Overview.html). *Proceedings of the Sixth International World Wide Web Conference (WWW6)*. Retrieved 10 January 2021.

6. Brin, Sergey; Page, Larry (1998). "The Anatomy of a Large-Scale Hypertextual Web Search Engine" (https://web.archive.org/web/20170713070157/http://ilpubs.stanford.edu:8090/361/1/1998-8.pdf) (PDF). *Proceedings of the Seventh International World Wide Web Conference (WWW7)*. Archived from the original (http://ilpubs.stanford.edu:8090/361/1/1998-8.pdf) (PDF) on 13 July 2017. Retrieved 10 January 2021.
7. Ken Harrenstien; Vic White (March 1982). *NICNAME/WHOIS* (https://www.rfc-editor.org/rfc/rfc812). doi:10.17487/RFC0812 (https://doi.org/10.17487%2FRFC0812). RFC 812 (https://datatracker.ietf.org/doc/html/rfc812). *Status Unknown.* Obsoleted by RFC 954 (https://www.rfc-editor.org/rfc/rfc954), 3912 (https://www.rfc-editor.org/rfc/rfc3912)
8. "Knowbot programming: System support for mobile agents" (http://www.cnri.reston.va.us/home/koe/iwooos-full.html). *cnri.reston.va.us*.
9. Deutsch, Peter (11 September 1990). "[next] An Internet archive server server (was about Lisp)" (https://groups.google.com/forum/#!msg/comp.archives/LWVA50W8BKk/wyRbF_lDc6cJ). *groups.google.com*. Retrieved 29 December 2017.
10. "World-Wide Web Servers" (http://www.w3.org/History/19921103-hypertext/hypertext/DataSources/WWW/Servers.html). W3C. Retrieved 14 May 2012.
11. "What's New! February 1994" (http://home.mcom.com/home/whatsnew/whats_new_0294.html). Mosaic Communications Corporation!. Retrieved 14 May 2012.
12. Search Engine Watch (September 2001). "Search Engines" (https://web.archive.org/web/20090413030108/http://www.internethistory.leidenuniv.nl/index.php3?c=7). *Internet History*. Netherlands: Universiteit Leiden. Archived from the original (http://www.internethistory.leidenuniv.nl/index.php3?c=7) on 13 April 2009.
13. "Archie" (https://www.pcmag.com/encyclopedia/term/archie). PCMag. Retrieved 20 September 2020.
14. Alexandra Samuel (21 February 2017). "Meet Alan Emtage, the Black Technologist Who Invented ARCHIE, the First Internet Search Engine" (https://daily.jstor.org/alan-emtage-first-internet-search-engine/). ITHAKA. Retrieved 20 September 2020.
15. loop news barbados. "Alan Emtage- a Barbadian you should know" (https://web.archive.org/web/20200923065914/http://www.loopnewsbarbados.com/content/alan-emtage-barbadian-you-should-know). loopnewsbarbados.com. Archived from the original (http://www.loopnewsbarbados.com/content/alan-emtage-barbadian-you-should-know) on 23 September 2020. Retrieved 21 September 2020.
16. Grandoni, Dino (16 April 2013). "Alan Emtage: The Man Who Invented The World's First Search Engine (But Didn't Patent It)" (https://www.huffingtonpost.co.uk/entry/alan-emtage-search-engine_n_2994090). *HuffPost*.
17. Oscar Nierstrasz (2 September 1993). "Searchable Catalog of WWW Resources (experimental)" (https://groups.google.com/group/comp.infosystems.www/browse_thread/thread/2176526a36dc8bd3/2718fd17812937ac?hl=en&lnk=gst&q=Oscar+Nierstrasz#2718fd17812937ac).
18. "Archive of NCSA what's new in December 1993 page" (https://web.archive.org/web/20010620073530/http://archive.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/old-whats-new/whats-new-1293.html). 20 June 2001. Archived from the original (http://archive.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/old-whats-new/whats-new-1293.html) on 20 June 2001. Retrieved 14 May 2012.
19. "What is first mover?" (https://searchcio.techtarget.com/definition/first-mover). *SearchCIO*. TechTarget. September 2005. Retrieved 5 September 2019.
20. Oppitz, Marcus; Tomsu, Peter (2017). *Inventing the Cloud Century: How Cloudiness Keeps Changing Our Life, Economy and Technology* (https://books.google.com/books?id=vrEvDwAAQBAJ&pg=PA238). Springer. p. 238. ISBN 9783319611617.
21. "Yahoo! Search" (https://web.archive.org/web/19961128070718/http://www.yahoo.com/search.html). *Yahoo!*. 28 November 1996. Archived from the original (https://www.yahoo.com/search.html) on 28 November 1996. Retrieved 5 September 2019.

22. Greenberg, Andy, "The Man Who's Beating Google" (https://www.forbes.com/forbes/2009/1005/technology-baidu-robin-li-man-whos-beating-google.html), *Forbes* magazine, October 5, 2009
23. Yanhong Li, "Toward a Qualitative Search Engine", *IEEE Internet Computing*, vol. 2, no. 4, pp. 24–29, July/Aug. 1998, doi:10.1109/4236.707687 (https://doi.org/10.1109%2F4236.707687)
24. "About: RankDex" (http://www.rankdex.com/about.html), *rankdex.com*
25. USPTO, "Hypertext Document Retrieval System and Method" (https://patents.google.com/patent/US5920859), US Patent number: 5920859, Inventor: Yanhong Li, Filing date: Feb 5, 1997, Issue date: Jul 6, 1999
26. "Baidu Vs Google: The Twins Of Search Compared" (https://fourweekmba.com/baidu-vs-google/). *FourWeekMBA*. 18 September 2018. Retrieved 16 June 2019.
27. Altucher, James (18 March 2011). "10 Unusual Things About Google" (https://www.forbes.com/sites/jamesaltucher/2011/03/18/10-unusual-things-about-google-also-the-worst-vc-decision-i-ever-made/). *Forbes*. Retrieved 16 June 2019.
28. "Method for node ranking in a linked database" (https://patents.google.com/patent/US6285999). Google Patents. Archived (https://web.archive.org/web/20151015185034/http://www.google.com/patents/US6285999) from the original on 15 October 2015. Retrieved 19 October 2015.
29. "Yahoo! And Netscape Ink International Distribution Deal" (https://web.archive.org/web/20131116112021/http://files.shareholder.com/downloads/YHOO/701084386x0x27155/9a3b5ed8-9e84-4cba-a1e5-77a3dc606566/YHOO_News_1997_7_8_General.pdf) (PDF). Archived from the original (http://files.shareholder.com/downloads/YHOO/701084386x0x27155/9a3b5ed8-9e84-4cba-a1e5-77a3dc606566/YHOO_News_1997_7_8_General.pdf) (PDF) on 16 November 2013. Retrieved 12 August 2009.
30. "Browser Deals Push Netscape Stock Up 7.8%" (https://www.latimes.com/archives/la-xpm-1996-04-01-fi-53780-story.html). *Los Angeles Times*. 1 April 1996.
31. Oremus, Will (14 October 2013). "Google's Big Break" (https://slate.com/business/2013/10/googles-big-break-how-bill-gross-goto-com-inspired-the-adwords-business-model.html). *Slate*. ISSN 1091-2339 (https://search.worldcat.org/issn/1091-2339). Retrieved 12 March 2025.
32. Jay (19 February 1998). "GoTo: The Forgotten Search Engine" (https://thehistoryoftheweb.com/goto-forgotten-search-engine/). *The History of the Web*. Retrieved 12 March 2025.
33. Gandal, Neil (2001). "The dynamics of competition in the internet search engine market" (http://www.escholarship.org/uc/item/0h17g08v). *International Journal of Industrial Organization*. **19** (7): 1103–1117. doi:10.1016/S0167-7187(01)00065-0 (https://doi.org/10.1016%2FS0167-7187%2801%2900065-0). ISSN 0167-7187 (https://search.worldcat.org/issn/0167-7187).
34. "Our history in depth" (https://web.archive.org/web/20121101210037/https://www.google.com/about/company/history/). Archived from the original (https://www.google.com/about/company/history/) on 1 November 2012. Retrieved 31 October 2012.
35. "Brave Search" (https://search.brave.com/help/brave-search-crawler).
36. "Definition – search engine" (https://www.techtarget.com/whatis/definition/search-engine). *Techtarget*. Retrieved 1 June 2023.
37. Jawadekar, Waman S (2011), "8. Knowledge Management: Tools and Technology" (https://books.google.com/books?id=XmGx4J9daUMC&pg=PA278), *Knowledge Management: Text & Cases*, New Delhi: Tata McGraw-Hill Education Private Ltd, p. 278, ISBN 978-0-07-07-0086-4, retrieved 23 November 2012
38. Dasgupta, Anirban; Ghosh, Arpita; Kumar, Ravi; Olston, Christopher; Pandey, Sandeep; and Tomkins, Andrew. *The Discoverability of the Web*. http://www.arpitaghosh.com/papers/discoverability.pdf

39. Jansen, B. J., Spink, A., and Saracevic, T. 2000. Real life, real users, and real needs: A study and analysis of user queries on the web (https://faculty.ist.psu.edu/jjansen/academic/pubs/jansen_real_life_real_users_and_real_needs.pdf). Information Processing & Management. 36(2), 207–227.
40. Chitu, Alex (30 August 2007). "Easy Way to Find Recent Web Pages" (http://googlesystem.blogspot.com/2007/08/easy-way-to-find-recent-web-pages.html). *Google Operating System*. Retrieved 22 February 2015.
41. "how search engine works?" (http://globalforumonline.com/detail/how-does-search-engine-works/). GFO. Retrieved 26 June 2018.
42. "What Is Local SEO & Why Local Search Is Important" (https://www.searchenginejournal.com/local-seo/what-is-local-seo-why-local-search-is-important/). *Search Engine Journal*. Retrieved 26 April 2020.
43. "Search Engine Market Share Worldwide" (https://www.similarweb.com/engines/). *Similarweb Top search engines*. Retrieved 19 February 2024.
44. Kerr, Dara (2 May 2024). "U.S. v. Google: As landmark 'monopoly power' trial closes, here's what to look for" (https://www.npr.org/2024/05/02/1248152695/google-doj-monopoly-trial-antitrust-closing-arguments). *NPR*.
45. "Search Engine Market Share Russia - June 2024" (https://gs.statcounter.com/search-engine-market-share/all/russian-federation). *StatCounter*. Retrieved 30 July 2025.
46. "Google ends China search redirect" (https://www.bbc.com/news/10423341). *BBC News*. 29 June 2010. Retrieved 30 July 2025.
47. "Google's China dilemma: to stay or to go" (https://www.reuters.com). *Reuters*. March 2023. Retrieved 30 July 2025.
48. "Search Engine Market Share China - June 2024" (https://gs.statcounter.com/search-engine-market-share/all/china). *StatCounter*. Retrieved 30 July 2025.
49. "Search Engine Market Share South Korea - June 2024" (https://gs.statcounter.com/search-engine-market-share/all/south-korea). *StatCounter*. Retrieved 30 July 2025.
50. "Search Engine Market Share Japan - June 2024" (https://gs.statcounter.com/search-engine-market-share/all/japan). *StatCounter*. Retrieved 30 July 2025.
51. "Search Engine Market Share Taiwan - June 2024" (https://gs.statcounter.com/search-engine-market-share/all/taiwan). *StatCounter*. Retrieved 30 July 2025.
52. Segev, El (2010). Google and the Digital Divide: The Biases of Online Knowledge (https://books.google.com/books?id=_Y9wAgAAQBAJ&q=bias), Oxford: Chandos Publishing.
53. Vaughan, Liwen; Mike Thelwall (2004). "Search engine coverage bias: evidence and possible causes". *Information Processing & Management*. **40** (4): 693–707. CiteSeerX 10.1.1.65.5130 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.65.5130). doi:10.1016/S0306-4573(03)00063-3 (https://doi.org/10.1016%2FS0306-4573%2803%2900063-3). S2CID 18977861 (https://api.semanticscholar.org/CorpusID:18977861).
54. Jansen, B. J. and Rieh, S. (2010) The Seventeen Theoretical Constructs of Information Searching and Information Retrieval (https://faculty.ist.psu.edu/jjansen/academic/jansen_theoretical_constructs.pdf). Journal of the American Society for Information Sciences and Technology. 61(8), 1517–1534.
55. Berkman Center for Internet & Society (2002), "Replacement of Google with Alternative Search Systems in China: Documentation and Screen Shots" (http://cyber.law.harvard.edu/filtering/china/google-replacements/), Harvard Law School.
56. Introna, Lucas; Helen Nissenbaum (2000). "Shaping the Web: Why the Politics of Search Engines Matters". *The Information Society*. **16** (3): 169–185. CiteSeerX 10.1.1.24.8051 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.24.8051). doi:10.1080/01972240050133634 (https://doi.org/10.1080%2F01972240050133634). S2CID 2111039 (https://api.semanticscholar.org/CorpusID:2111039).

57. Hillis, Ken; Petit, Michael; Jarrett, Kylie (12 October 2012). *Google and the Culture of Search* (https://archive.org/details/googlecultureofs0000hill). Routledge. ISBN 9781136933066.
58. Reilly, P. (1 January 2008). " 'Googling' Terrorists: Are Northern Irish Terrorists Visible on Internet Search Engines?". In Spink, Prof Dr Amanda; Zimmer, Michael (eds.). *Web Search*. Information Science and Knowledge Management. Vol. 14. Springer Berlin Heidelberg. pp. 151–175. Bibcode:2008wsis.book..151R (https://ui.adsabs.harvard.edu/abs/2008wsis.book..151R). doi:10.1007/978-3-540-75829-7_10 (https://doi.org/10.1007%2F978-3-540-75829-7_10). ISBN 978-3-540-75828-0. S2CID 84831583 (https://api.semanticscholar.org/CorpusID:84831583).
59. Hiroko Tabuchi, "How Climate Change Deniers Rise to the Top in Google Searches (https://www.nytimes.com/2017/12/29/climate/google-search-climate-change.html)", The New York Times, Dec. 29, 2017. Retrieved November 14, 2018.
60. Ballatore, A (2015). "Google chemtrails: A methodology to analyze topic representation in search engines" (https://doi.org/10.5210%2Ffm.v20i7.5597). *First Monday*. **20** (7). doi:10.5210/fm.v20i7.5597 (https://doi.org/10.5210%2Ffm.v20i7.5597).
61. Pariser, Eli (2011). *The filter bubble : what the Internet is hiding from you*. New York: Penguin Press. ISBN 978-1-59420-300-8. OCLC 682892628 (https://search.worldcat.org/oclc/682892628).
62. O'Hara, K. (1 July 2014). "In Worship of an Echo" (https://doi.org/10.1109%2FMIC.2014.71). *IEEE Internet Computing*. **18** (4): 79–83. Bibcode:2014IIC....18d..79O (https://ui.adsabs.harvard.edu/abs/2014IIC....18d..79O). doi:10.1109/MIC.2014.71 (https://doi.org/10.1109%2FMIC.2014.71). ISSN 1089-7801 (https://search.worldcat.org/issn/1089-7801). S2CID 37860225 (https://api.semanticscholar.org/CorpusID:37860225).
63. Bruns, Axel (29 November 2019). "Filter bubble" (https://policyreview.info/node/1426). *Internet Policy Review*. **8** (4). doi:10.14763/2019.4.1426 (https://doi.org/10.14763%2F2019.4.1426). hdl:10419/214088 (https://hdl.handle.net/10419%2F214088). ISSN 2197-6775 (https://search.worldcat.org/issn/2197-6775). S2CID 211483210 (https://api.semanticscholar.org/CorpusID:211483210).
64. Haim, Mario; Graefe, Andreas; Brosius, Hans-Bernd (2018). "Burst of the Filter Bubble?" (https://doi.org/10.1080%2F21670811.2017.1338145). *Digital Journalism*. **6** (3): 330–343. doi:10.1080/21670811.2017.1338145 (https://doi.org/10.1080%2F21670811.2017.1338145). ISSN 2167-0811 (https://search.worldcat.org/issn/2167-0811). S2CID 168906316 (https://api.semanticscholar.org/CorpusID:168906316).
65. Nechushtai, Efrat; Lewis, Seth C. (2019). "What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations" (https://linkinghub.elsevier.com/retrieve/pii/S0747563218303650). *Computers in Human Behavior*. **90**: 298–307. doi:10.1016/j.chb.2018.07.043 (https://doi.org/10.1016%2Fj.chb.2018.07.043). S2CID 53774351 (https://api.semanticscholar.org/CorpusID:53774351).
66. "New Islam-approved search engine for Muslims" (https://web.archive.org/web/20130712023215/http://news.msn.com/science-technology/new-islam-approved-search-engine-for-muslims). News.msn.com. Archived from the original (https://news.msn.com/science-technology/new-islam-approved-search-engine-for-muslims) on 12 July 2013. Retrieved 11 July 2013.
67. "Jewogle - FAQ" (https://web.archive.org/web/20190207015631/http://www.jewogle.com/faq/). Archived from the original (http://www.jewogle.com/faq/) on 7 February 2019. Retrieved 6 February 2019.
68. "Halalgoogling: Muslims Get Their Own "sin free" Google; Should Christians Have Christian Google? - Christian Blog" (https://web.archive.org/web/20140913102516/http://allchristiannews.com/halalgoogling-muslims-get-their-own-sin-free-google-should-christians-have-christian-google/). *Christian Blog*. 25 July 2013. Archived from the original (http://allchristiannews.com/halalgoogling-muslims-get-their-own-sin-free-google-should-christians-have-christian-google/) on 13 September 2014. Retrieved 13 September 2014.

69. Schwartz, Barry (29 October 2012). "Google: Search Engine Submission Services Can Be Harmful" (https://www.seroundtable.com/search-engine-submission-google-15906.html). *Search Engine Roundtable*. Retrieved 4 April 2016.
70. Heymann, Paul; Koutrika, Georgia; Garcia-Molina, Hector (12 February 2008). "Can Social Bookmarking Improve Web Search?" (http://dbpubs.stanford.edu:8090/pub/2008-2). *First ACM International Conference on Web Search and Data Mining*. Retrieved 12 March 2008.
71. Priti Srinivas Sajja; Rajendra Akerkar (2012). *Intelligent technologies for web applications* (https://books.google.com/books?id=HqXxoWK7tucC&q=the+University+of+Nevada+System+Computing+Services+group+developed+Veronica.&pg=PA87). Boca Raton: CRC Press. p. 87. ISBN 978-1-4398-7162-1. Retrieved 3 June 2014.
72. "A History of Search Engines" (http://www.wiley.com/legacy/compbooks/sonnenreich/history.html). Wiley. Retrieved 1 June 2014.
73. Priti Srinivas Sajja; Rajendra Akerkar (2012). *Intelligent technologies for web applications* (https://books.google.com/books?id=HqXxoWK7tucC&q=the+University+of+Nevada+System+Computing+Services+group+developed+Veronica.&pg=PA87). Boca Raton: CRC Press. p. 86. ISBN 978-1-4398-7162-1. Retrieved 3 June 2014.
74. "The Major Search Engines" (https://web.archive.org/web/20140605052335/http://www.pccua.edu/kholland/major_search_engines.htm). 21 January 2014. Archived from the original (http://www.pccua.edu/kholland/major_search_engines.htm) on 5 June 2014. Retrieved 1 June 2014.
75. Jansen, B. J., Spink, A., Bateman, J., and Saracevic, T. 1998. Real life information retrieval: A study of user queries on the web (https://faculty.ist.psu.edu/jjansen/academic/jansen_sigir_forum.pdf). SIGIR Forum, 32(1), 5 -17.
76. Priti Srinivas Sajja; Rajendra Akerkar (2012). *Intelligent technologies for web applications* (https://books.google.com/books?id=HqXxoWK7tucC&q=the+University+of+Nevada+System+Computing+Services+group+developed+Veronica.&pg=PA87). Boca Raton: CRC Press. p. 85. ISBN 978-1-4398-7162-1. Retrieved 3 June 2014.
77. Li, Jingjing; Larsen, Kai; Abbasi, Ahmed (1 December 2020). "TheoryOn: A Design Framework and System for Unlocking Behavioral Knowledge Through Ontology Learning" (https://misq.org/theoryon-a-design-framework-and-system-for-unlocking-behavioral-knowledge-through-ontology-learning.html). *MIS Quarterly*. **44** (4): 1733–1772. doi:10.25300/MISQ/2020/15323 (https://doi.org/10.25300%2FMISQ%2F2020%2F15323). S2CID 219401379 (https://api.semanticscholar.org/CorpusID:219401379).

Further reading

- Steve Lawrence; C. Lee Giles (1999). "Accessibility of information on the web" (https://doi.org/10.1038%2F21987). *Nature*. **400** (6740): 107–9. Bibcode:1999Natur.400..107L (https://ui.adsabs.harvard.edu/abs/1999Natur.400..107L). doi:10.1038/21987 (https://doi.org/10.1038%2F21987). PMID 10428673 (https://pubmed.ncbi.nlm.nih.gov/10428673). S2CID 4347646 (https://api.semanticscholar.org/CorpusID:4347646).
- Bing Liu (2007), *Web Data Mining: Exploring Hyperlinks, Contents and Usage Data (http://www.cs.uic.edu/~liub/WebMiningBook.html).* Springer,ISBN 3-540-37881-2
- Bar-Ilan, J. (2004). The use of Web search engines in information science research. ARIST, 38, 231–288.
- Levene, Mark (2005). *An Introduction to Search Engines and Web Navigation*. Pearson.
- Hock, Randolph (2007). *The Extreme Searcher's Handbook*.ISBN 978-0-910965-76-7
- Javed Mostafa (February 2005). "Seeking Better Web Searches". *Scientific American*. **292** (2): 66–73. Bibcode:2005SciAm.292b..66M (https://ui.adsabs.harvard.edu/abs/2005SciAm.292b..66M). doi:10.1038/scientificamerican0205-66 (https://doi.org/10.1038%2Fscientificamerican0205-66).

- Ross, Nancy; Wolfram, Dietmar (2000). "End user searching on the Internet: An analysis of term pair topics submitted to the Excite search engine". *Journal of the American Society for Information Science*. **51** (10): 949–958. doi:10.1002/1097-4571(2000)51:10<949::AID-ASI70>3.0.CO;2-5 (https://doi.org/10.1002%2F1097-4571%282000%2951%3A10%3C949%3A%3AAID-ASI70%3E3.0.CO%3B2-5).
- Xie, M.; et al. (1998). "Quality dimensions of Internet search engines". *Journal of Information Science*. **24** (5): 365–372. doi:10.1177/016555159802400509 (https://doi.org/10.1177%2F016555159802400509). S2CID 34686531 (https://api.semanticscholar.org/CorpusID:34686531).
- *Information Retrieval: Implementing and Evaluating Search Engines* (https://web.archive.org/web/20201005195805/http://www.ir.uwaterloo.ca/book/). MIT Press. 2010. Archived from the original (http://www.ir.uwaterloo.ca/book/) on 5 October 2020. Retrieved 7 August 2010.
- Yeo, ShinJoung. (2023) *Behind the Search Box: Google and the Global Internet Industry* (U of Illinois Press, 2023) ISBN 0-252-08712-7 JSTOR 4116455 (https://www.jstor.org/stable/4116455)

External links

Retrieved from "https://en.wikipedia.org/w/index.php?title=Search_engine&oldid=1336076491"